New methods of interpretation using marginal effects for nonlinear models

Scott Long

Departments of Sociology and Statistics
Indiana University

EUSMEX 2016: Mexican Stata Users Group
Mayo 18, 2016

Road map for talk

Goals
1. Present new methods of interpretation using marginal effects
2. Show how to implement these methods with Stata

Outline
1. Statistical background
   - Binary logit model
   - Standard definitions of marginal effects
   - Generalizations of marginal effects
2. Stata commands
   - Estimation using factor notation, storing estimates, and gsem
   - Post-estimation using margins and lincom
   - SPost13’s m* commands
3. Example modeling the occurrence of diabetes

Logit model

Nonlinear in probability
\[ \pi(x) = \frac{\exp(x'\beta)}{1 + \exp(x'\beta)} = \Lambda(x'\beta) \]

Marginal effect: additive change in probability for change in \( x_k \) holding other variables at specific values

Multiplicative in odds
\[ \Omega(x) = \frac{\pi(x)}{1 - \pi(x)} = \exp(x'\beta) \]

Odds ratio: multiplicative change in \( \Omega(x) \) for change in \( x_k \) holding other variables constant

Marginal effects

1. Marginal change: instantaneous rate of change in \( \pi(x) \)
2. Discrete change: change in \( \pi(x) \) for discrete change in \( x \)

Definition of discrete change

1. Variable \( x_k \) changes from start to end
2. The remaining \( x \)'s are held constant at specific values \( x = x^* \)
3. Discrete change for \( x_k \)
   \[ DC(x_k) = \frac{\Delta \pi(x)}{\Delta x_k (\text{start} \rightarrow \text{end})} = \pi(x_k = \text{end}, x = x^*) - \pi(x_k = \text{start}, x = x^*) \]
4. Interpretation
   For a change in \( x_k \) from start to end, the probability changes by \( DC(x_k) \), holding other variables at the specified values.
Examples of discrete change

1. DC conditional on the specific values $x^*$

$$\frac{\Delta \pi(x = x^*)}{\Delta x_k(0 \rightarrow 1)} = \pi(x_k = 1, x = x^*) - \pi(x_k = 0, x = x^*)$$

2. DC conditional on the observed values for observation $i$

$$\frac{\Delta \pi(x = x_i)}{\Delta x_k(x_k \rightarrow x_k + 1)} = \pi(x_k = x_k + 1, x = x_i) - \pi(x_k = x_k, x = x_i)$$

The challenge of summarizing the effect of $x_k$

Since the value of $\Delta \pi / \Delta x_k$ depends on where it is evaluated, how do you summarize the effect?

Common summary measures of discrete change

**DC at the mean: change at the center of the data**

$$DCM(x_k) = \frac{\Delta \pi(x = x_k)}{\Delta x_k(start \rightarrow end)} = \pi(x_k = \text{end}, x) - \pi(x_k = \text{start}, x)$$

For someone who is average on all variables, increasing $x_k$ from start to end changes the probability by $DCM(x_k)$.

**Average DC: average change in estimation sample**

$$ADC(x_k) = \frac{1}{N} \sum_{i=1}^{N} \frac{\Delta \pi(x = x_i)}{\Delta x_k(start \rightarrow end)}$$

On average, increasing $x_k$ from start to end changes the probability by $ADC(x_k)$.

Variations in computing discrete change

**Conditional and average change**

- Conditional on specific values
- Averaged in the estimation sample
- Averaged in a subsample

**Type of change**

- Additive change
- Proportional change
- Changes as a function of $x$’s
- Change of a component of a multiplicative measure

**Number of variables changed**

- One variable
- Two or more mathematically linked variables
- Two or more substantively related variables

Stata installation, data, and do-files

1. Examples use Stata 14.1, but most things can be done with Stata 13
2. Requires the spost13ado package
3. Examples and slides available with search eusmex

Stata commands

1. Fitting logit model with factor syntax
   $$\text{logit depvar i.var c.var c.var1#c.var2}$$
2. Regression estimates are stored and restored
   $$\text{estimates store ModelName}$$
   $$\text{estimates restore ModelName}$$
3. margins estimates predictions from current regression results
4. margins, post stores these predictions allowing lincom to estimate functions of predictions
5. mchange, mtable, mgen and mlincm are SPost wrappers
Modeling diabetes

1. Cross-section data from Health and Retirement Survey
2. Outcome is self-report of diabetes
   2.1 Small changes are substantively important
   2.2 Since changes can be statistically significant since N=16,071
3. Road map for examples
   3.1 Compute standard measures of change to explain commands
   3.2 Extend these commands to compute complex types of effects
   3.3 Illustrate testing equality of effects within and across models

---

1Steve Heeringa generously provided the data used in Applied Survey Data Analysis (Heeringa et al., 2010). Complex sampling is not used in my analyses.

---

Two primary model specifications

1. Model Mbmi includes the BMI index
   logit diabetes c.bmi ///
   i.white c.age##c.age i.female i.hsdegree
   estimates store Mbmi
2. Model Mwt includes height and weight
   logit diabetes c.weight c.height ///
   i.white c.age##c.age i.female i.hsdegree
   estimates store Mwt
3. The estimates are...

---

Odds ratios and p-values tell us little

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mbmi</th>
<th>Mwt</th>
</tr>
</thead>
<tbody>
<tr>
<td>bmi</td>
<td>1.1046*</td>
<td></td>
</tr>
<tr>
<td>weight</td>
<td>1.0165*</td>
<td></td>
</tr>
<tr>
<td>height</td>
<td>0.9299*</td>
<td></td>
</tr>
<tr>
<td>white</td>
<td>0.5412*</td>
<td>0.5313*</td>
</tr>
<tr>
<td>age</td>
<td>1.3091*</td>
<td>1.3093*</td>
</tr>
<tr>
<td>c.age#c.age</td>
<td>0.9983*</td>
<td>0.9983*</td>
</tr>
<tr>
<td>female</td>
<td>0.7848*</td>
<td>0.8743#</td>
</tr>
<tr>
<td>hsdegree</td>
<td>0.7191*</td>
<td>0.7067*</td>
</tr>
<tr>
<td>_cons</td>
<td>0.0000*</td>
<td>0.0001*</td>
</tr>
<tr>
<td>bic</td>
<td>14991.26</td>
<td>14982.03</td>
</tr>
</tbody>
</table>

Note: # significant at .05 level; * at the .001 level.

---

Average discrete change

1. mchange is a useful first step after fitting a model
   . estimates restore Mbmi
   . mchange, amount(sd) // compute average discrete change
   logit: Changes in Pr(y) | Number of obs = 16071
<table>
<thead>
<tr>
<th>Change</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>bmi</td>
<td>+0.097</td>
</tr>
<tr>
<td>white</td>
<td></td>
</tr>
<tr>
<td>White vs Non-white</td>
<td>-0.099</td>
</tr>
</tbody>
</table>
   (output omitted)
2. Interpretation
   Increasing BMI by one standard deviation on average increases the probability of diabetes .097.
   On average, the probability of diabetes is .099 less for white respondents than non-white respondents.
3. Where did these numbers come from?

---

Tool: margins, at( ...) and atmeans

1. By default,
   1.1 margins computes prediction for every observation
   1.2 Then the predictions are averaged
2. Options allow predictions at “counterfactual” values of variables
3. Average prediction assuming everyone is white
   margins, at(white=1)
4. Two average predictions under two conditions
   margins, at(white=1) at(white=0)
5. Conditional prediction if white with means for other variables
   margins, at(white=1) atmeans
ADC for binary $x_k$: ADC(white)

1. ADC(white) is the difference in average probabilities
   \[
   \text{ADC} = \frac{1}{N} \sum_i \pi(\text{white} = 1, x = x_i) - \frac{1}{N} \sum_i \pi(\text{white} = 0, x = x_i)
   \]

2. margins computes the two averages
   \[
   \text{margins, at(white=0) at(white=1) post}
   \]

   Expression: \( \Pr(\text{diabetes}), \text{predict}() \)

   \begin{align*}
   &1._{\text{at}}: \text{white} \ = \ 0 \\
   &2._{\text{at}}: \text{white} \ = \ 1
   \end{align*}

   Delta-method
   \[
   \begin{array}{cccc}
   \text{Std. Err.} & \text{z} & \text{P>|z|} & \text{[95% Conf. Interval]} \\
   \hline
   \text{1.}_{\text{at}} & .2797806 & .0073107 & 38.27 & 0.000 & .265452 & .2941092 \\
   \text{2.}_{\text{at}} & .1805306 & .0034215 & 52.76 & 0.000 & .1738245 & .1872367 \\
   \end{array}
   \]

3. \(1._{\text{at}}\) is the average treating everyone as nonwhite
   \[
   1._{\text{at}} = \frac{1}{N} \sum_i \pi(\text{white} = 0, x = x_i)
   \]

4. \(2._{\text{at}}\) is the average treating everyone as white

Tool: mlincom simplifies lincom

1. \text{mlincom} requires column names from e(b) that can be complex
   \[
   \text{mlincom} \ (_b[2._{\text{at}} #1.\text{white}] - _b[1._{\text{at}} #1.\text{white}]) \ \text{///} \ \text{lincom} \ (_b[2._{\text{at}} #0.\text{white}] - _b[1._{\text{at}} #0.\text{white}])
   \]

2. \text{mlincom} uses column numbers in e(b) or rows in margins output
   \text{mlincom} (4-2) - (3-1)

ADC for continuous $x_k$: ADC(bmi + sd)

1. Compute probabilities at observed bmi and observed + sd
   \[
   . \text{quietly sum bmi} \\
   . \text{local sd = r(sd)} \\
   . \text{margins, at(bmi = gen(bmi)) at(bmi = gen(bmi + \text{sd})) post}
   \]
   Expression: \( \Pr(\text{diabetes}), \text{predict}() \)

   \begin{align*}
   &1._{\text{at}}: \text{bmi} = \text{bmi} \\
   &2._{\text{at}}: \text{bmi} = \text{bmi} + 5.770835041238605
   \end{align*}

   Delta-method
   \[
   \begin{array}{cccc}
   \text{Std. Err.} & \text{z} & \text{P>|z|} & \text{[95% Conf. Interval]} \\
   \hline
   \text{1.}_{\text{at}} & .2047166 & .0030338 & 67.48 & 0.000 & .1987704 & .2106627 \\
   \text{2.}_{\text{at}} & .3017056 & .005199 & 58.03 & 0.000 & .2915159 & .3118954 \\
   \end{array}
   \]

2. ADC(bmi + sd)
   \[
   \text{mlincom} \ 2._{\text{at}} - 1._{\text{at}} \text{ stats(all)}
   \]
   \[
   \begin{array}{cccccc}
   \text{lincom} & \text{se} & \text{zvalue} & \text{pvalue} & \text{ll} & \text{ul} \\
   \hline
   1 & 0.097 & 0.044 & 27.208 & 0.000 & 0.090 & 0.104 \\
   \end{array}
   \]

On average, increasing BMI by one standard deviation, about 6 points, increases the probability of diabetes by .097 \((p < .001)\).

Tool: margins, at( varnm = generate(exp) )

1. \text{margins, at( varnm = generate(exp) )} is a powerful, nearly undocumented option that generates values for making predictions

2. Trivially, average prediction at observed values of bmi
   \text{margins, at(bmi = gen(bmi))}

3. Average prediction at observed values plus 1
   \text{margins, at(bmi = gen(bmi+1))}

4. Two average predictions
   \text{margins, at(bmi = gen(bmi)) at(bmi = gen(bmi+1))}

5. Average at observed plus standard deviation
   \begin{enumerate}
   \item quietly sum bmi
   \item local sd = r(sd)
   \item margins, at(bmi = gen(bmi + \text{sd}))
   \end{enumerate}

Tool: mtable wrapper for margins

1. \text{margins} output is complete, not compact

2. \text{mtable} executes \text{margins} and simplifies the output (and more)
   \begin{itemize}
   \item \text{mtable, commands} lists the \text{margins} commands used
   \item \text{mtable, detail} shows \text{margins} output and \text{mtable} output
   \end{itemize}
DCM for continuous $x_k$: DCM(bmi + sd)

Discrete change at the mean

1. Let bmi increase from mean(bmi) to mean(bmi) + sd(bmi)
   
   ```
   . qui sum bmi
   . local mn = r(mean)
   . local mnplus = r(mean) + r(sd)
   ```

2. Option `atmeans` holds other variables at their means

   ```
   . margins, atmeans at(bmi = `mn´) at(bmi = `mnplus´) post
   ```

   Expression : Pr(diabetes), predict()

   ```
   1._at : bmi = 27.89787
   0.white = .2284239 (mean)
   1.white = .7715761 (mean)
   age = 69.29276 (mean)
   0.female = .4315226 (mean)
   1.female = .5684774 (mean)
   0.hsdegree = .2375086 (mean)
   1.hsdegree = .7624914 (mean)
   ```

   ```
   2._at : bmi = 33.6687
   0.white = .2284239 (mean)
   1.white = .7715761 (mean)
   age = 69.29276 (mean)
   0.female = .4315226 (mean)
   1.female = .5684774 (mean)
   0.hsdegree = .2375086 (mean)
   1.hsdegree = .7624914 (mean)
   ```

   2. Alternatively, `mtable` runs `margins` and reformats the results

   ```
   . mtable, atmeans at(bmi = `mn´) at(bmi = `mnplus´) post
   ```

   Expression: Pr(diabetes), predict()

   ```
   bmi Pr(y)
   1 27.9 0.210
   2 33.7 0.320
   ```

   Specified values of covariates

   ```
<table>
<thead>
<tr>
<th></th>
<th>1. white</th>
<th>age</th>
<th>female</th>
<th>hsdegree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>.772</td>
<td>.63</td>
<td>.568</td>
<td>.762</td>
</tr>
<tr>
<td>Specified</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.722</td>
<td></td>
<td>.693</td>
<td>.568</td>
</tr>
</tbody>
</table>
   ```

   For an average person, increasing BMI by one standard deviation increases the probability of diabetes by .111 ($p < .001$).

DCM for continuous $x_k$: DCM(bmi + sd)

2. Alternatively, `mtable` runs `margins` and reformats the results

   ```
   . mtable, atmeans at(bmi = `mn´) at(bmi = `mnplus´) post
   ```

   Expression: Pr(diabetes), predict()

   ```
   bmi Pr(y)
   1 27.9 0.210
   2 33.7 0.320
   ```

   Specified values of covariates

   ```
<table>
<thead>
<tr>
<th></th>
<th>1. white</th>
<th>age</th>
<th>female</th>
<th>hsdegree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current</td>
<td>.772</td>
<td>.63</td>
<td>.568</td>
<td>.762</td>
</tr>
<tr>
<td>Specified</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.722</td>
<td></td>
<td>.693</td>
<td>.568</td>
</tr>
</tbody>
</table>
   ```

   For an average person, increasing BMI by one standard deviation increases the probability of diabetes by .111 ($p < .001$).

Generalized measures of discrete change

1. `mchange` makes the above computations automatically

2. I did it the hard way to illustrate powerful commands

3. Now these commands are used for some interesting things

Proportional change in $x_k$:

1. Body mass can be measured with height and weight

   ```
   . logit diabetes c.weight c.height ///
   . i.white c.age##c.age i.female i.hsdegree
   ```

2. ADC(weight + 25) increases weight by 25 pounds for everyone

3. Increasing weight 25 pound is a
   - 25% increase from 100 pounds
   - 14% increase from average weight
   - 8% increase from 300 pounds

4. Is the effect of a percentage increase in weight more meaningful than an additive increase?

5. First, compute ADC(weight+25)...
Proportional change in $x_k$: ADC(weight*1.14)

2. A simple change computes ADC(weight * 1.14)
   - estimates restore Mwt
   - at(weight = gen(weight)) at(weight = gen(weight * 1.14)) post
   - Expression: Pr(diabetes), predict()
   - \[
   \begin{array}{c|c|c|c|c}
   \text{Pr(y)} & 1 & 0.205 & 2 & 0.273 \\
   \end{array}
   \]
   - mlincom 2 - 1, rowname(ADC pct) add
   - lincom pvalue ll ul
   - \[
   \begin{array}{l|c|c|c}
   \text{ADC add} & 0.067 & 0.000 & 0.062 & 0.071 \\
   \text{ADC pct} & 0.068 & 0.000 & 0.063 & 0.073 \\
   \end{array}
   \]

3. The effects are deceptively similar

Discrete change with polynomials

1. With polynomials multiple variables must change together
2. For example,
   \[
   \frac{\Delta\pi(x)}{\Delta age(50 \rightarrow 60)} = \pi(age=60, agesq=60^2) - \pi(age=50, agesq=50^2)
   \]
3. This can be done two ways
   3.1 The easy way with factor syntax
   3.2 The hard way with at(... = gen(...))

Discrete change with polynomials

1. With $x$ and $x^2$ only values on the blue curve are mathematically possible
2. Changes in the probability reflect linked changes in $x$ and $x^2$

Discrete change with polynomials

1. When $c.age$ changes, margins automatically changes $c.age#c.age$

Tool: factor notation for polynomials

Without factor notation
1. Generate age-squared
   - generate agesq = age * age
2. Model specification
   - logit diabetes c.age c.agesq ...

With factor notation
1. Model specification where $c.$ is necessary
   - logit diabetes c.age##c.age ...
2. $c.age##c.age$ does three things
   2.1 Adds $c.age$ to the model
   2.2 Creates $c.age#c.age \equiv c.age*c.age$
   2.3 Adds $c.age#c.age$ to the model
3. When $c.age$ changes, margins automatically changes $c.age#c.age$
Discrete change with age & age^2

Correct ADC with factor notation

1. age and age#age automatically change together
   - age and age#age automatically change together
   - logit diabetes c.age##c.age c.bmi i.white i.female i.hdegree, or
   - post
   - mtable, at(age = gen(age)) at(age = gen(age+10)) post
   - Expression: Pr(diabetes), predict()
   - Pr(y)
   - 1 0.205
   - 2 0.223
   - mlincom 2 - 1

2. Interpretation
   On average, a ten-year increase in age increases the probability of diabetes by .02 (p < .001).

Discrete change with associated variables

1. Age and age-squared are mathematically linked
2. Other variables could be substantively associated
3. Example: To examine the effect of cultural capital on health, change all cultural assets together, not a single asset
4. Example: Are “larger people” (taller people with the same body mass) more likely to have diabetes?
   - ▶ Use height to predict weight
   - ▶ Use margins, at(...=gen()) to change height and weight together

This example illustrates the power of margins, at(...=gen())

Associated variables: ADC(height, weight)

1. Regress weight on height and height squared
   - regress weight c.height##c.height, noci
   - R-squared = 0.2575
   - weight Coef. Std. Err. t P>|t|
   - height -6.338708 1.61073 -3.94 0.000
   - c.height#c.height .0855799 .0120867 7.08 0.000
   - _cons 217.5991 53.5548 4.06 0.000

2. Save estimates
   - scalar b0 = _b[_cons]
   - scalar b1 = _b[height]
   - scalar b2 = _b[c.height#c.height]

3. Use at(gen(...)) to predicts weight assuming a 6” change in height
   - at( height = gen(height) /// observed height
   - weight = gen(weight) ) /// observed weight
   - height = gen(height+6) /// +6 inches
   - weight = gen(b0 + b1* (height+6) /// +estimated weight
   - + b2*((height+6)^2)) ) //

   Expression: Pr(diabetes), predict()
   - Pr(y)
   - 1 0.205
   - 2 0.223
   - mlincom 2 - 1

4. Interpretation
   There is no evidence that being physically larger without greater body mass contributes to the incidence of diabetes.

Distribution of effects

1. ADC and DCM are common summary measures of change
2. Each uses the mean to summarize a distribution
3. ADC: average discrete change
   \[ \text{ADC}(x_i) = \frac{1}{N} \sum_j \frac{\Delta \pi(x_i|x = x_j)}{\Delta(x_i|x = x_j)} \]
4. DCM: discrete change at the mean
   \[ \text{DCM}(x_i) = \frac{\Delta \pi}{\Delta(x_i|x = x)} \]
   where \( x_k = \frac{1}{N} \sum_j x_k \)
5. Hypothetical data shows why means can be misleading
**Distribution of effects: ADC and DCM**

Hypothetical data

6. **Does age affect diabetes** since ADC(age) and DCM(age) are near 0?

![Distribution of effects: ADC and DCM](image)

**Undocumented Tool: margins, generate()**

1. margins, gen(*stub*) creates variables with predictions for each observation (help margins generate)
2. For example, to save probabilities for 16,071 cases and average them
   - margins, gen(Prob)
   - Predictive margins
   - Expression : Pr(diabetes), predict()
   - Number of obs = 16,071

| Margin | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|--------|-----------|-------|-----|---------------------|
| _cons  | .2047166  | .0030316 | 67.53 | 0.000  | .1987747 .2106584 |

. sum Prob1 // matches margins estimate

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob1</td>
<td>16,071</td>
<td>.2047166</td>
<td>.1229016</td>
<td>.0123593</td>
<td>.9067207</td>
</tr>
</tbody>
</table>

**Distribution of effects: ADC(age)**

1. To evaluate ADC(age) look at the distribution of DC(agei)
2. Create a variable with the DC for each observation
   - 1) margins, generate(PRage) ///
   - 2) at(age = gen(age)) at(age = gen(age+10))
   - 3) gen DCage10 = PRage2 - PRage1
   - 4) lab var DCage10 "DC for 10 year increase in age"

**Distribution of effects: ADC(weight)**

1. Now consider ADC(weight+25) and ADC(weight*1.14)
   - 1) mtable, gen(PRadd) at(weight=gen(weight)) at(weight=gen(weight+25)) post
   - 2) generate DCadd = PRadd2 - PRadd1
   - 3) lab var DCadd "DC for 25 pound increase"
   - 4) mtable, gen(PRpct) at(weight=gen(weight)) at(weight=gen(weight*1.14)) post
   - 5) generate DCpct = PRpct2 - PRpct1
   - 6) lab var DCpct "DC for 14 percent increase in weight"
2. The changes have very different distributions

**Distribution of effects: ADC(weight)**

3. While the ADCs are close, effects for individuals can differ greatly
Distribution of effects: limitations of summaries

1. ADC and DCM are more useful than odds ratios!
2. In nonlinear models, any summary measures can be misleading
3. The distribution of effects is valuable for assessing effects and is simple with margins, generate()
   - Long and Freese (2014) show how do this without the gen() option
4. For age, multiple DCRs are more useful than ADC or DCM

Comparing DCRs

1. Are the DCR(age) significantly different at different ages?

Comparing DCRs at different ages

2. Compute probabilities at 4 ages with other variables at means
   - mtable, at(age=(60(10)90)) post
   - Expression: Pr(diabetes), predict()

3. DCRs at different ages
   - mlincom 2-1, clear rowname(DCR60)
   - mlincom 3-2, add rowname(DCR70)
   - mlincom 4-3, add rowname(DCR80)

4. Test differences in DCRs
   - mlincom (2-1) - (3-2), add rowname(DCR60 - DCR70)
   - mlincom (2-1) - (4-3), add rowname(DCR60 - DCR80)
   - mlincom (3-2) - (4-3), add rowname(DCR70 - DCR80)

5. Summarizing
   - mlincom, twidth(14)

Comparing ADCs for two variables

1. ADC(race) and ADC(bmi+sd) have similar size, but different signs
   - quietly sum bmi
   - local sd = r(sd)
   - margins, at(white=(0 1)) at(bmi = gen(bmi)) at(bmi = gen(bmi + `sd´)) post

2. Can you justify saying the effects have the same size?
3. To test equality they must be estimated simultaneously

Comparing ADC(white) and ADC(bmi)

1. Simultaneously compute components of ADC(white) and ADC(bmi)
Comparing ADC(white) and ADC(bmi)

4. Compute effects and test equality
   . qui mlincom (2-1), rowname(ADC white) clear
   . qui mlincom (4-3), rowname(ADC bmi) add
   . mlincom (2-1) + (4-3), rowname(Sum of ADCs) add

<table>
<thead>
<tr>
<th>lincom</th>
<th>pvalue</th>
<th>ll</th>
<th>ul</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADC female</td>
<td>-0.099</td>
<td>0.000</td>
<td>-0.115</td>
</tr>
<tr>
<td>ADC bmi</td>
<td>0.097</td>
<td>0.000</td>
<td>0.090</td>
</tr>
<tr>
<td>Sum of ADCs</td>
<td>-0.002</td>
<td>0.809</td>
<td>-0.021</td>
</tr>
</tbody>
</table>

5. Conclusion
   The health cost of being non-white is equivalent to a standard deviation increase in body mass (p > .80).

Comparing ADC across subsamples

1. An ADC is typically averaged over the estimation sample
2. By averaging within groups, we can examine effects for different groups
   - Is the average effect of BMI the same for whites and non-whites?
3. This requires margins, over()

Tool: margins, over()

1. By default, margins averages over all observations
2. Averages on subsamples are possible with if and over()
3. Averaging for the non-white subsample
   margins if white==0, ///
   at(bmi = gen(bmi)) at(bmi = gen(bmi+`sd'))
4. For the white subsample
   margins if white==1, ///
   at(bmi = gen(bmi)) at(bmi = gen(bmi+`sd'))
5. For both subsamples simultaneously
   margins, over(white) ///
   at(bmi = gen(bmi)) at(bmi = gen(bmi+`sd'))

Comparing ADC(bmi) by race

2. Computing ADC(bmi) by group
   . qui mlincom 4-2, clear rowname(ADC white: ADC bmi)
   . mlincom 3-1, add rowname(Non-white: ADC bmi)

<table>
<thead>
<tr>
<th>lincom</th>
<th>pvalue</th>
<th>ll</th>
<th>ul</th>
</tr>
</thead>
<tbody>
<tr>
<td>White: ADC bmi</td>
<td>0.090</td>
<td>0.000</td>
<td>0.083</td>
</tr>
<tr>
<td>Non-white: ADC bmi</td>
<td>0.121</td>
<td>0.000</td>
<td>0.112</td>
</tr>
</tbody>
</table>

3. A second difference compares effects for the groups
   . mlincom (4-2) - (3-1), rowname(Difference: ADC bmi)

<table>
<thead>
<tr>
<th>lincom</th>
<th>pvalue</th>
<th>ll</th>
<th>ul</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference: ADC bmi</td>
<td>-0.030</td>
<td>0.000</td>
<td>-0.034</td>
</tr>
</tbody>
</table>

4. Interpretation
   The average effect of BMI is significantly larger for non-whites than whites (p < .001).

Comparing ADC(bmi) by race

2. Computing ADC(bmi) by group
   . qui mlincom (2-1), rowname(ADC white: ADC bmi)
   . qui mlincom (4-3), rowname(ADC bmi) add
   . mlincom (2-1) + (4-3), rowname(Sum of ADCs) add

<table>
<thead>
<tr>
<th>lincom</th>
<th>pvalue</th>
<th>ll</th>
<th>ul</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADC female</td>
<td>-0.099</td>
<td>0.000</td>
<td>-0.115</td>
</tr>
<tr>
<td>ADC bmi</td>
<td>0.097</td>
<td>0.000</td>
<td>0.090</td>
</tr>
<tr>
<td>Sum of ADCs</td>
<td>-0.002</td>
<td>0.809</td>
<td>-0.021</td>
</tr>
</tbody>
</table>

Comparing ADCs across models

1. Does the effect of a variable change with model specification?
2. Tool: margins, dydx(female) computes DC(female) since \( i.female \)
3. Computing ADC(female) for two models
   . qui logit diabetes c.bmi i.female i.white c.age##c.age i.hsdegree
   . qui mtable, dydx(female) rowname(ADC(female) with Mbmi) clear
   . qui logit diabetes c.weight c.height i.female i.white c.age##c.age i.hsdegree
   . mtable, dydx(female) rowname(ADC(female) with Mwt) below

   Expression: \( \text{Pr(diabetes)} \), predict()

   | Delta-method | Margin | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
   |--------------|--------|-----------|-----|--------|----------------------|
   | at#White     |        |           |     |        |                      |
   | 1Non-white   | .3097249 | .0072773 | 42.56 | 0.000   | .2954616 - .3239881 |
   | 1White       | .173629 | .0032892 | 52.79 | 0.000   | .1671824 - .1800757 |
   | 2Non-white   | .4302294 | .009226  | 46.63 | 0.000   | .4121468 - .448312  |
   | 2White       | .2636564 | .0054903 | 48.02 | 0.000   | .2528955 - .2744172 |

4. To test if they are equal, the effects must be estimated simultaneously

\( d \text{ Pr(y)} \)

<table>
<thead>
<tr>
<th>ADC(female) with Mbmi</th>
<th>-0.036</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADC(female) with Mwt</td>
<td>-0.020</td>
</tr>
</tbody>
</table>
**Tool:** simultaneous model estimation with `gsem`

1. `gsem` simultaneously fits multiple GLM models

2. The obvious approach does not work since

   ```
   gsem ///
   (diabetes <- c.bmi i.female i.white c.age##c.age i.hsdegree, logit) ///
   (diabetes <- c.weight c.height i.female i.white c.age##c.age i.hsdegree, logit)
   ///
   , vce(robust)
   ```

   is interpreted as a single model

3. The solution is to create cloned outcomes for each model

   ```
   clonevar lhsbmi = diabetes // outcome for bmi model
   clonevar lhswt = diabetes // outcome for weight height model
   ```

### Comparing ADC(female) across models

2. Estimate ADC(female) for both models simultaneously

   ```
   qui restore Mgsem
   margins, dydx(female) post
   ```

   Average marginal effects
   Model VCE : Robust
dy/dx w.r.t. : 1.female
1._predict : Predicted mean (Respondent has diabetes?), predict(pr
outcome(outcome(lhsbmi))
2._predict : Predicted mean (Respondent has diabetes?), predict(pr
outcome(lhswt))

   Delta-method
dy/dx Std. Err. z P>|z| [95% Conf. Interval]

   1.female
   _predict 1 -0.0360559 .0061773 -5.84 0.000 -.0481631 -.0239487
   2 -0.0199213 .0089687 -2.22 0.026 -.0374997 -.0023429

   Note: dy/dx for factor levels is the discrete change from the base level.

### Comparing ADC(female) across models

3. Test if ADC(female) is the same in both models

   ```
   mlincom 1-2, stats(all)
   ```

   lincom se zvalue pvalue ll ul
   |
   1 -0.016 0.006 -2.526 0.012 -0.029 -0.004

   4. Interpretation

   **The effect of being female is significantly larger when body mass is measured with the BMI index then when height and weight are used** $(p < .02)$. 

### Comparing effects across models: summary

1. Jointly estimating models with `gsem` and computing effects with `margins` is a general approach for comparing effects across models (Mize et al., 2009)

2. `gsem`
   1. Fits the GLM class of models, but does not fit non-GLM models
   2. `margins` is slow (grumble, grumble)

3. `suest`
   1. Fits a much wider class of models
   2. `margins` is fast, but much harder to use (grumble, grumble)

4. `suest` and `gsem` produce identical results

5. Specialized commands like `kxb` (Kohler et al., 2011) are available

### Comparing groups

**Linear regression**

1. Coefficients differ by group such as $\beta^W_{\text{female}}$ and $\beta^N_{\text{female}}$

2. Analysis focuses on Chow tests such as $H_0: \beta^N_{\text{female}} = \beta^W_{\text{female}}$

**Logit and probit**

1. Coefficients differ by group such as $\beta^W_{\text{female}}$ and $\beta^N_{\text{female}}$

2. The coefficients combines
   1. The effect of $x_k$ which can differ by group
   2. The variance of the error which can differ by group

3. Since regression coefficients are identified to a scale factor, Chow-type tests of $H_0: \beta^N = \beta^W$ are invalid (Allison, 1999)

4. Probabilities and marginal effects are identified (Long, 2009)
Comparing groups: outcomes and effects

Group differences can be examined two ways

1. Differences in probabilities

\[ H_0: \pi_W(x = x^*) = \pi_N(x = x^*) \]

Is the probability of diabetes the same for white and non-white respondents who have the same characteristics?

2. Differences in marginal effects

\[ H_0: \frac{\Delta \pi_W}{\Delta x_k} = \frac{\Delta \pi_N}{\Delta x_k} \]

Is the effect of \( x_k \) the same for whites and non-whites?

3. These dimensions of difference are shown in the next graph

Comparing groups: model estimation

1. Factor syntax allows coefficients to differ by white

\[
\text{logit diabetes ibn.white ///}
\]

\[
\text{ibn.white##(i.female i.hsdegree c.age##c.age c.bmi), nocon}
\]

2. This is equivalent to simultaneously estimating

\[
\text{logit diabetes i.female i.hsdegree c.age##c.age c.bmi if white==1}
\]

\[
\text{logit diabetes i.female i.hsdegree c.age##c.age c.bmi if white==0}
\]

3. Resulting in these estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>Whites</th>
<th>NonWhites</th>
<th>female</th>
<th>0.713 1.024</th>
<th>0.000 0.755</th>
<th>odds ratios</th>
<th>p-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsdegree</td>
<td>0.706 0.743</td>
<td>0.000 0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>1.278 1.369</td>
<td>0.000 0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comparing groups: probabilities by age

1. Compute DC(white) at various ages

\[
\text{- mtable, dydx(white) at(age=(55(10)85)) atmeans stats(est p)}
\]

Expression: \( \text{Pr(diabetes), predict()} \)

<table>
<thead>
<tr>
<th>age</th>
<th>d Pr(y)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>55</td>
<td>-0.078</td>
<td>0.000</td>
</tr>
<tr>
<td>65</td>
<td>-0.124</td>
<td>0.000</td>
</tr>
<tr>
<td>75</td>
<td>-0.129</td>
<td>0.000</td>
</tr>
<tr>
<td>85</td>
<td>-0.092</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Specified values of covariates

| 0. 1. 1. 1. |
| white white female hsdegree bmi |
| Current .228 .772 .568 .762 27.9 |

2. Example of interpretation

The probability of diabetes is significantly larger for 55 year-old non-whites than whites who are average on other characteristics (\( p < .01 \)).

3. Graphically we can show effects at multiple ages

Note: these plots can be computed with mgen or marginsplot
Comparing groups: ADC or DCM?

Comparing ADCs
1. Differences in ADCs are determined by both
   1.1 Differences in the probability curves
   1.2 Differences in distribution of variables

Comparing DCRs
1. DCRs show differences in probability curves at a specific location
2. DCRs do not depend on the distribution of variables

Which to use?
1. The answer depends on what you want to know?

Comparing groups: ADC(bmi + 5)

1. To compute ADC(bmi + 5) by race
   - .stable over(white) gen(bmi) gen(bmi+5) at(bmi = gen(bmi)) at(bmi = gen(bmi+5)) post
   - .mtable, over(white) at(bmi = gen(bmi)) at(bmi = gen(bmi+5)) post
   - .pr diabetes, predict()

2. Conclusion
   The average effects of BMI are not significantly different for whites and non-whites (p=.83).

Comparing groups: DCR(age + 10)

1. Since ADC(age) might not be useful due to nonlinearity, we compare DCR(age+10) at different ages
   1.1 Other variables are held at sample means
   1.2 Group specific means could be used (Long and Freese, 2014)

2. For example, DCR(age + 10) at 55
   - .stable at(age=55 white=0 1) at(age=65 white=0 1) atmeans post
   - .mtable, at(age=55 white=(0 1)) at(age=65 white=(0 1)) atmeans post
   - .mlincom 3-1, rowname(DC nonwhite) stats(est p) clear
   - .mlincom 4-2, rowname(DC white) stats(est p) add
   - .mlincom (4-2) - (3-1), rowname(Dif at 55) stats(est p) add

3. And so on, with the following results

* Decomposing an effect

1. The BMI index measures relative weight or body mass
   \[ BMI = \frac{weight}{height^2} = 703 \times \frac{weight_{lb}}{height_{in}^2} \]

2. Question 1: With BMI in the model, what is the effect of weight?
   - Why do this? DC(weight) is clearer to patients than DC(bmi)

3. Question 2: Does DC(weight) depend on how body mass is measured?

4. To answer these questions think of BMI as an interaction
   \[ BMI = 703 \times weight \times height^{-1} \times height^{-1} \]

Decomposing BMI: BMI as an interaction

1. Create components of BMI
   - .generate heightinv = 1/height
   - .generate S = 703

2. These models are identical
   - .logit diabetes c.bmi i.white c.age##c.age i.female i.hsdegree
   - .estimates store Mbmi
   - .logit diabetes c.S##c.weight##c.heightinv##c.heightinv
     i.white c.age##c.age i.female i.hsdegree
   - .estimates store MbmiFV

3. The estimates are identical
   - Variable MbmiFV Mbmi
   - \[ S = 703 \]
   - \[ \text{odds ratio for BMI} \]
   - \[ \text{odds ratio for BMI} \]
   - \[ White White \]
   - \[ .000 0.000 \]
Decomposing BMI: ADC(weight)

4. margins with factor syntax makes the rest trivial

5. ADC(weight) in MbmiFV changes only weight
   . qui estimates restore MbmiFV
   . mchange weight, amount(sd) delta(25)
   . logit: Changes in Pr(y) | Number of obs = 16071
   Expression: Pr(diabetes), predict(pr)

<table>
<thead>
<tr>
<th>Change</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>weight</td>
<td>+25</td>
</tr>
</tbody>
</table>

6. ADC(weight) in Mwt is slightly larger
   . qui estimates restore Mwt
   . mchange weight, amount(sd) delta(25)
   . logit: Changes in Pr(y) | Number of obs = 16071
   Expression: Pr(diabetes), predict(pr)

<table>
<thead>
<tr>
<th>Change</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>weight</td>
<td>+25</td>
</tr>
</tbody>
</table>

Decomposing an effect: summary

1. Factor variables and margins make the difficult decompositions trivial

2. Factor syntax understands interactions in model specifications

3. margins in turn understands interactions and handles the messy details

Comparing ADC(weight) across models

1. To compare ADC(weight) requires joint estimation
   . clonevar lshbmi = diabetes
   . clonevar lhswt = diabetes
   . gof ///
   > (lshbmi < c.s#c.weight#c.heightinv#c.heightinv ///
   > 1.white c.age##c.age 1.female 1.hsdegree, logit) ///
   > (lhswt < c.weight c.height ///
   > 1.white c.age##c.age 1.female 1.hsdegree, logit) ///
   > , vce(robust)

   Generalized structural equation model Number of obs = 16,071
   Response : lshbmi
   Family : Bernoulli
   Link : logit
   Response : lhswt
   Family : Bernoulli
   Link : logit
   Log pseudolikelihood = -14914.007
   (output omitted)

2. Computing the average predictions for both equations
   . margins, at(weight=gen(weight)) at(weight=gen(weight+25))
   Predictive margins Number of obs = 16,071
   Model VCE : Robust
   1._predict : Predicted mean (Diabetes?), predict(pr outcome(lhsbmi))
   2._predict : Predicted mean (Diabetes?), predict(pr outcome(lhswt))
   1._at : weight = weight
   2._at : weight = weight+25

   Delta-method
   Margin Std. Err. z P>|z| [95% Conf. Interval]
   1._predict#_at
   1 1 .2047166 .0030419 67.30 0.000 .1987546 .2106786
   2 1 .2701404 .0044591 60.58 0.000 .2614007 .27888
   1 2 .2701404 .0044591 60.58 0.000 .2614007 .27888
   2 2 .271305 .0044054 61.58 0.000 .2626705 .2799394

Comparing ADC(weight) in two models

3. ADC(weight) for each model and their difference
   . qui mlincom 2-1, rowname(Mbmi ADC) clear
   . qui mlincom 4-3, rowname(Mwt ADC) add
   . mlincom (4-3) - (2-1), rowname(Difference)

<table>
<thead>
<tr>
<th>Lincom</th>
<th>p-value</th>
<th>ul</th>
<th>ll</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mbmi ADC</td>
<td>0.065</td>
<td>0.000</td>
<td>0.061</td>
</tr>
<tr>
<td>Mwt ADC</td>
<td>0.067</td>
<td>0.000</td>
<td>0.062</td>
</tr>
<tr>
<td>Difference</td>
<td>0.001</td>
<td>0.029</td>
<td>0.000</td>
</tr>
</tbody>
</table>

4. Conclusion
   The effect of weight on diabetes are nearly identical whether body mass is measured with BMI or with height and weight (p = .03).

Conclusions

Model interpretation and Stata

1. Too often interpretation ends with estimated coefficients

2. Interpretation using predictions is more informative
   - I think of regression coefficients as “nuisance parameters”

3. Methods of interpretation must be practical

4. margins makes hard things easy, very hard things merely hard

5. Hopefully, Stata 15 will make impossible things possible
Conclusions

Marginal effects are only one method of interpretation

1. Standard marginal effects are more useful than odds ratios
2. mchange allows marginal effects to be a routine part of analysis
3. Today’s talk shows how to customize marginal effects for the substantive application
4. However, marginal effects are not the only or best method of interpretation
5. Tables and plots are often valuable (Long and Freese, 2014)
6. The best interpretation is motivated by your substantive question

Thanks to many people

Thank you for listening

Collaborators Parts of this work were developed with Long Doan, Jeremy Freese, Trent Mize, and Sarah Mustillo. Jeff Pitblado and David Drukker provided valuable help. Mistakes are my own.

Relevant publications There is a large literature on marginal effects and interpreting models. Long and Freese (2014) include many citations. The references directly related to this presentation are given below.

Bibliography


